

## Abstract of CN1295295

The fast index polling structure includes a word index data part, a secondary index data part, and a word related content index data part. According to the user input word, the method looks up word data base and word content related data base and utilizes the positioning characteristics of the secondary index data part to speed looking-up, reduce looking-up number and raise looking-up effect.

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.<sup>7</sup>

G06F 17/30

## [12] 发明专利申请公开说明书

[21] 申请号 99123297.6

[43] 公开日 2001 年 5 月 16 日

[11] 公开号 CN 1295295A

[22] 申请日 1999.11.4 [21] 申请号 99123297.6  
[71] 申请人 英业达集团(西安)信息技术有限公司  
地址 710000 陕西省西安市高新技术产业开发区  
F-2B 号楼 2 楼  
[72] 发明人 陈淮琰 李 玲

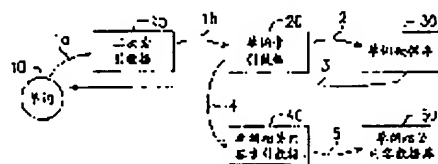
[74] 专利代理机构 柳沈知识产权律师事务所  
代理人 马 莹

权利要求书 3 页 说明书 11 页 附图页数 7 页

[54] 发明名称 使用快速查询索引结构的电子字典单词  
查询方法

[57] 摘要

一种使用快速查询索引结构的电子字典的单词查询方法,该快速查询索引结构包括单词索引数据部,二次索引数据部以及单词相关内容索引数据部。该方法根据用户的输入单词,查询单词数据库和单词内容相关数据库,利用二次索引数据部的定位特性,加速查询速度,减少查询范围和次数,提高查询功能。



ISSN 1008-4274

1. 一种电子字典的单词查询方法，所述电子字典包括：

中央处理单元，为整个系统的控制中心，控制执行该方法的各个步骤；

5 输入设备，用于输入用户想要查询的字串；

显示设备，用于显示用户输入的字串和查询内容；

存储器，其中存放有单词数据库和单词相关内容数据库，所述单词数据库包括已排序的多个待查询单词并且所述单词相关内容数据库包括分别对应于所述待查询单词的多个单词相关内容，

10 其特征在于：

所述存储器中还存储有电子字典快速查询索引结构，包括：

单词索引数据部，用以储存对应于所述待查询单词的多个单词索引值，所述单词索引值用以记录对应的待查询单词在所述单词数据库中的位置信息；

15 二次索引数据部，用以储存分别对应于预定字母数的有序序列中的项目的多个二次索引值，所述二次索引值用以记录其开始字母与有序序列项目相同的待查询单词，其对应的单词索引值在所述单词索引数据部的位置信息；以及

20 单词相关内容索引数据部，用以储存对应于所述单词相关内容的多个单词相关内容索引值，其中所述单词相关内容索引值用以记录对应的单词相关内容在所述单词相关内容数据库中的位置信息，

该方法利用所述快速查询索引结构，根据用户的输入单词，查询所述单词数据库和所述单词内容相关数据库，其包括下列步骤：

25 中央处理单元根据用户从输入设备输入的单词中与所述有序序列的预定字母数相同的开始字母部分，从所述二次索引部取出一第一二次索引值；

中央处理单元从所述二次索引部取出所述第一二次索引值的下一个的第二二次索引值；

30 中央处理单元以所述第一二次索引值和所述第二二次索引值所对应的单词索引值为范围，从所述单词索引数据部找出对应于所述输入单词的单词索引值；以及

根据对应于所述输入单词的单词索引值，从所述单词相关内容索引数据



部中找出对应的单词相关内容索引值。

2. 如权利要求 1 所述的方法，其中所述快速查询索引结构中的所述二次索引数据部中，所述二次索引值系对应于两个字母的有序序列项目。

3. 如权利要求 1 所述的方法，其中在所述第一二次索引值和所述第二二次索引值所对应的单词索引值为范围中，利用连续折半查询方法，找出对应于所述输入单词的单词索引值。

4. 如权利要求 1 所述的单词查询方法，其中在所述二次索引数据部中，所述二次索引值系对应于两个字母的有序序列项目。

5. 一种电子字典的单词查询方法，所述电子字典包括：

10 中央处理单元，为整个系统的控制中心，控制执行该方法的各个步骤；  
输入设备，用于输入用户想要查询的字串；

显示设备，用于显示用户输入的字串和查询内容；

存储器，其中存放有单词数据库和单词相关内容数据库，所述单词数据库包括已排序的多个待查询单词并且所述单词相关内容数据库包括分别对应  
15 于所述待查询单词的多个单词相关内容，

其特征在于：

所述存储器中还存储有电子字典快速查询索引结构，包括：

单词索引数据部，用以储存对应于所述待查询单词的多个单词索引值，  
所述单词索引值用以记录对应的待查询单词在所述单词数据库中的位置信  
20 息；

二次索引数据部，用以储存分别对应于预定字母数的有序序列中的项目的  
多个二次索引值，所述二次索引值用以记录其开始字母与有序序列项目相  
同的待查询单词，其对应的单词索引值在所述单词索引数据部的位置信息；  
以及

25 单词相关内容索引数据部，用以储存对应于所述单词相关内容的多个  
单词相关内容索引值，其中所述单词相关内容索引值用以记录对应的单词相  
关内容在所述单词相关内容数据库中的位置信息，

该方法利用所述快速查询索引结构，根据用户的输入单词，动态地查  
询所述单词数据库和所述单词相关内容数据库，当所述输入单词的字母数不  
30 大于所述有序序列的预定字母数时，执行下列步骤：

中央处理单元根据所述输入单词，从所述二次索引部取出一第一二次

索引值；

中央处理单元从所述二次索引部取出所述第一二次索引值的下一个的第二二次索引值；

5 设定所述第一二次索引值所对应的第一单词索引值和所述第二二次索引值所对应的第二单词索引值为一查询范围；以及

根据所述第一二次索引值所对应的第一单词索引值，从所述单词相关内容索引数据部中找出对应的单词相关内容索引值；

当所述输入单词的字母数大于所述有序序列的预定字母数，执行下列步骤：

10 以所述第一单词索引值和所述第二单词索引值为查询范围，从所述单词索引数据部找出对应于所述输入单词的第三单词索引值；

以所述第三单词索引值取代所述第一单词索引值；以及

根据取代后的所述第一单词索引值，从所述单词相关内容索引数据部中找出对应的单词相关内容索引值。

15 6. 如权利要求 5 所述的单词查询方法，其中在所述第一单词索引值和所述第二单词索引值为范围中，利用连续折半查询方法，找出对应于所述输入单词的第三单词索引值。

7. 如权利要求 5 所述的单词查询方法，其中在所述二次索引数据部中，所述二次索引值对应于两个字母的有序序列项目。

# 使用快速查询索引结构的 电子字典单词查询方法

5

本发明涉及一种使用快速查询索引结构的电子字典单词查询方法，特别涉及利用二次索引数据和单词索引数据所建立的索引结构来快速查询单词，并且能够随着用户逐次输入单词各字母的同时，动态地快速查询数据的单词查询方法。

10

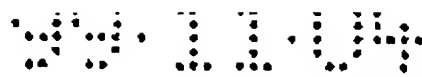
对于电脑上的电子字典来说，一般在用户输入单词的过程中，会利用输入单词的部分内容(例如前几个字母)来查询单词的相关内容，例如释义，同时还必须快速地查询出所输入单词部分内容的后续单词。换言之，每当用户输入待查询单词的任何一个字母时，电子字典就必须执行一遍查询动作，因此查询动作的快慢势必影响到用户对于电子字典整体性能的感觉。

15

图1表示传统电子字典执行动态单词查询的索引数据结构以及查询处理步骤的数据流图。其中，10表示用户所输入的单词，其可能是完整的待查询单词亦或是待查询单词的部分输入内容。20是单词索引数据，30则是单词数据库，图2表示单词数据库30的结构，其中，单词数据库30储存所有可查询的单词，而单词索引数据20则是用来储存指出某个单词在单词数据库30的储存位置的索引值。举例来说，如果单词数据库30中的所有可查询单词以连续方式储存，则单词“a”在单词索引数据20中对应项目的索引值设为0，单词“aardvark”在单词索引数据20中对应项目的索引值设为1(前一单词为单字母)，单词“aardwolf”在单词索引数据20中对应项目的索引值设为9(前两个单词合计有9个字母)，其余依此类推。40是单词相关内容索引数据，50是单词相关内容数据库。单词相关内容数据库50是用来储存所有可查询单词所对应到的内容，包括释义，文法等等信息；而单词相关内容

20

25

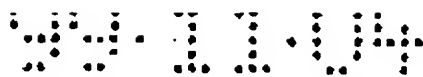


索引数据 40 则是用来储存指出某个单词相关内容在单词相关内容数据库 50 的储存位置的索引值。

5 从于单词索引数据 20、单词数据库 30、单词相关内容索引数据 40 和单词相关内容数据库 50 为固定数据,因此在一般电子字典中是利用只读存储器(mask ROM)加以储存。另外,对于包含数万条或是数十万条单词的电子字典而言,单词数据库 30 和单词相关内容数据库 50 的数据量非常的大。为了节省储存空间,降低成本,一般会利用数据压缩的方式处理这些数据库。考虑到解压缩效率,一般会采用区块压缩的方式来减少每次解压缩所要处理的数据量。

10 以下说明现有技术单词查询处理流程。当用户输入单词 10 后,可以先利用折半查询方法来查询出对应的单词索引值。首先透过步骤 1,在单词索引数据 20 中找出第一个和最后一个单词索引值,进行折半查询来找出真正对应此单词 10 的期望索引值。接着透过步骤 2,从单词数据库 30 中找出对应此期望索引值的期望单词,如果单词数据库为压缩数据时,则在步骤 2 必须  
15 同时进行解压缩的处理。透过步骤 3,从所找出的期望单词与单词 10 进行比较,如果两者相同,则表示已经找到对应的单词索引值;如果两者不同,则更新折半查询范围,重新执行步骤 1 到步骤 3,直到找到真正的对应单词索引值为止。当找到对应单词 10 的单词索引值后,就可以透过步骤 4,从单词相关内容索引数据 40 找出对应的单词相关内容索引值;再利用所找到的单词  
20 相关内容索引值,透过步骤 5,从单词相关内容数据库 50 中找出对应的相关内容。

接着以一实例来说明现有技术中的单词查询过程。图 3~图 6 表示使用电子字典输入查询单词“schema”的过程中,电子字典显示查询结果的示意图。在图 3 中,用户输入查询单词的部分字母“s”,系统会将单词索引数据  
25 20 的第一项索引值和最后一项索引值之间设定为查询范围,进行连续性的折半查询;当系统查询到单词数据库 30 内的单词“S”之后,则结束连续性



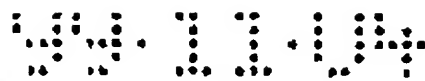
折半查询，再透过单词相关内容索引数据 40 和单词相关内容数据库 50，找出其相关内容以便显示。接着在图 4 至图 6 的处理部分也与所述方式一致，用户依序输入“sc”、“sch”、“schema”，系统则利用折半查询方式，连续地进行查询/对比的动作，直到确认查询目标为止。

5 从上述说明的单词查询流程可以发现，整个处理最耗时的部分是找出对应的单词索引值，亦即步骤 1 至步骤 3 的循环处理。特别是对于具有十几万或数十万的大字典而言，重复查询的次数非常庞大。如果数据是利用压缩方式储存的，在步骤 2 的解压缩动作也会耗费相当多的时间，这也正是传统查询方法速度较慢的原因。

10 有鉴于此，本发明的主要目的在于提供一种使用电子字典快速查询索引结构的单词查询方法，能够简化用户在逐字输入单词过程中所需要的查询动作，以提高电子字典使用的效率。

根据上述的目的，本发明提出一种使用电子字典快速查询索引结构的单词查询方法，该电子字典包括：中央处理单元，为整个系统的控制中心，控制执行本发明方法的各个步骤；输入设备，用于输入用户想要查询的字串；  
15 显示设备，用于显示用户输入的字串和查询内容；存储器，其中存放有单词数据库和单词相关内容数据库，该单词数据库包括已排序的多个待查询单词并且该单词相关内容数据库包括分别对应于该待查询单词的多个单词相关内容，该存储器中还存储有电子字典快速查询索引结构，包括：(1)单词索引  
20 数据部，用来储存对应于待查询单词的多个个单词索引值，单词索引值是用以记录对应的待查询单词在单词数据库中的位置信息；(2)二次索引数据部，用以储存多个个二次索引值，其分别对应于具有预定字母数的有序序列中的各项目，而二次索引值则是用来记录具有与有序序列项目相同的开始字母的待查询单词，其对应的单词索引值在单词索引数据部中的位置信息；以及(3)  
25 单词相关内容索引数据部，用以储存对应于单词相关内容的多个个单词相关内容索引值，单词相关内容索引值则是用来记录对应的单词相关内容在单词





相关内容数据库中的位置信息。

当利用上述索引数据结构来进行单词查询时，中央处理器先根据输入单词中与有序序列的预定字母数相同的开始字母部分，从二次索引部取出第一二次索引值，接着再从二次索引部取出第一二次索引值的下一个的第二二次索引值。接着再以第一二次索引值和第二二次索引值所对应的单词索引值为范围，从单词索引数据部找出对应于输入单词的单词索引值。最后便可以根据对应于输入单词的单词索引值，从单词相关内容索引数据部中找出对应的单词相关内容索引值。

按照本发明方法的另一方面，当利用上述索引数据结构来进行单词查询时，中央处理器可根据用户所输入的字母长度来做不同的处理。当输入单词的字母数不大于有序序列的预定字母数时，则先根据输入单词，从二次索引部取出第一二次索引值，从二次索引部取出所述第一二次索引值的下一个的第二二次索引值，设定第一二次索引值所对应的第一单词索引值和第二二次索引值所对应的第二单词索引值为一查询范围，接着根据第一二次索引值所对应的第一单词索引值，从单词相关内容索引数据部中找出对应的单词相关内容索引值。另外，当输入单词的字母数大于有序序列的预定字母数时，便可以利用第一单词索引值和第二单词索引值为查询范围，从单词索引数据部找出对应于输入单词的第三单词索引值。将查询到的第三单词索引值取代第一单词索引值，便可以利用取代后的第一单词索引值，从单词相关内容索引数据部中找出对应的单词相关内容索引值。此部分处理则随用户的输入持续进行。

为使本发明的上述目的、特征和优点能更明显易懂，下文特举一优选实施例，并配合附图，作详细说明如下：

图1表示现有技术电子字典执行动态单词查询的索引数据结构以及查询处理步骤的数据流程图。

图2表示一般单词数据库结构的示意图。

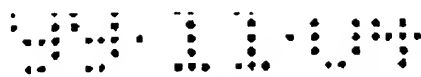


图 3 ~ 图 6 表示使用电子字典输入查询单词 “ schema ” 的过程中，电子字典显示查询结果的示意图。

图 7 表示本发明实施例中执行动态单词查询所使用的索引数据结构以及查询处理步骤的数据流图。

5 图 8 表示本发明实施例中使用的二次索引数据结构的示意图。

图 9 表示本发明实施例中使用的二次索引数据、单词索引数据以及单词数据库之间关联性的示意图。

图 10 表示本发明实施例执行动态单词查询的处理流程图。

图 11 是实施本发明方法的电子字典的硬件组成配置图。

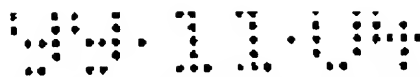
10 符号说明：

10 ~ 单词； 15 ~ 二次索引数据； 20 ~ 单词索引数据； 30 ~ 单词数据库； 40 ~ 单词相关内容索引数据； 50 ~ 单词相关内容数据库。

本发明的电子字典使用快速查询索引结构的单词查询方法，主要是透过二次索引数据来减少查询单词索引数据时所需要进行的循环数，同时可以减少执行单词数据库解压缩处理的次数，提高查询的效率。特别是适合于电子字典动态查询的应用上。以下配合图式，详细说明本发明的实施例。

图 11 是实施本发明方法的电子字典的硬件组成配置图。实施本发明方法的电子字典的硬件配置包括：中央处理单元 1，为整个系统的控制中心，控制执行本发明方法的各个步骤；输入设备 3，用于输入用户想要查询的字串；显示设备 2，用于显示用户输入的字串和查询内容；存储器 4，其中存放有单词数据库和单词相关内容数据库，该单词数据库包括已排序的多个待查询单词并且该单词相关内容数据库包括分别对应于该待查询单词的多个单词相关内容，该存储器中还存储有电子字典快速查询索引结构。

图 7 表示本实施例中执行动态单词查询的索引数据结构以及查询处理步骤的数据流图。在图 7 中，除了系统必备的单词数据库 30 和单词相关内容数据库之外，索引数据结构中包括了二次索引数据 15、单词索引数据 20 和单



词相关内容索引数据 40。其中单词索引数据 20 和单词相关内容索引数据 40 与现有技术中所述者相同，此处不再赘述，以下则详细说明二次索引数据 20 的结构。

二次索引数据 20 是用来储存二次索引值，其分别对应于具有预定字母数的有序序列中的各项目。有序序列是指长度不超过预定字母数而依照字母顺序(类似一般字典中的排序方式)排列而成的序列。在本实施例中，有序序列的预定字母数为 2 个，因此其有序序列可以表示为(a、aa、ab、ac、…、b、ba、bb、bc、…c、ca、cb、cc、…zy、zz)，其项目总数为  $26 \times 27 = 702$ 。二次索引值则是用来记录开始字母与有序序列项目相同的待查询单词，其对应的单词索引值在单词索引数据部 20 的位置信息。

图 8 表示本实施例中二次索引数据结构的示意图。如图所示，二次索引数据 15 包括彼此对应的有序序列部 15b 以及二次索引值部 15a。必须说明的是，二次索引数据 15 实际置于存储器中者只有二次索引值部 15a，有序序列部 15b 则为虚拟对应数据。当需要找出某个有序序列项目在二次索引数据 20 中所对应的二次索引值，可以透过以下公式决定：

$$(\text{第一字母顺序编号} - 1) \times 27 + \text{第二字母顺序编号}(1)$$

举例来说，有序序列项目“sc”所对应的二次索引值系置于二次索引数据 20 中的第  $(19-1) \times 27 + 3 = 489$  个位置中。另外，实际的二次索引值是用来指向以对应有序序列项目为首的第一个待查询单词的单词索引值位置。例如有序序列项目“sc”所对应的二次索引值为 47，表示在单词索引数据 20 中第 47 个位置的单词索引值所对应的待查询单词，是以“sc”为首并且在单词数据库中排序在最前面的待查询单词。

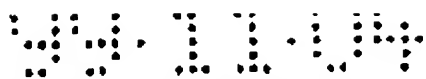
图 9 可以用来说明本实施例中二次索引数据 15、单词索引数据 20 以及单词数据库 30 之间关联性的示意图。如图所示，二次索引数据 15 中的二次索引值，是用来指向单词索引数据 20 中对应于其有序序列项的第一个单词的单词索引项；而单词索引数据 20 中的单词索引项则指向其对应单词在单词数



数据库 30 中的储存位置。例如，对应于有序序列项目 “aa” 的二次索引值为 1，指向单词索引数据 20 中的第 1 个单词索引值，而此单词索引值则指向以 “aa” 为开头的第一个单词 “aardvard” 的储存位置；对应于有序序列项目 “ab” 的二次索引值为 30，指向单词索引数据 20 中的第 30 个单词索引值，而此单词索引值则指向以 “ab” 为开头的第一个单词 “ab” 的储存位置；对应于有序序列项目 “ba” 的二次索引值为 183，指向单词索引数据 20 中的第 183 个单词索引值，而此单词索引值则指向以 “ab” 为开头的第一个单词 “backer” 的储存位置。

回到图 7，以下说明利用本实施例的索引数据结构进行单词查询的过程。首先，接收用户所输入的单词 10，在步骤 1a 中，根据单词 10 的开始字母部分(在本实施例中为两个字母)，利用上述的公式(1)计算出单词 10 所对应的二次索引值的储存位置；再透过此储存位置从二次索引数据 15 取出二次索引值 I1。同时，从二次索引数据 15 取出二次索引值 I1 的下一个二次索引值 I2。在步骤 1b 中，则根据在上一步骤中得到的二次索引值 I1 和二次索引值 I2，从单词索引数据 20 中找出对应的单词索引值 R1 和单词索引值 R2。在步骤 2 和步骤 3 中，则以单词索引值 R1 和单词索引值 R2 为范围，进行连续的折半查询/对比的处理，藉以从单词索引数据 20 找出对应于单词 10 的单词索引值。此折半查询/对比处理与现有技术方式相同，亦即以折半查询方式找出期望索引值，再从单词数据库 30 中找出对应于此期望索引值的期望单词(步骤 2)，期望单词与实际的单词 10 进行对比后(步骤 3)，则可以决定是否找到符合的单词索引值，如果尚未找到，则利用期望索引值来更新折半查询范围，反复进行直到找出为止。当找出对应于单词 10 的单词索引值之后，便可以透过单词相关内容索引数据 40 找出其对应的单词相关内容索引值(步骤 4)，再从单词相关内容数据库 50 找出其对应的单词相关内容(步骤 5)。

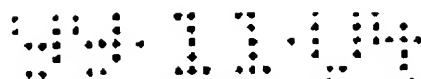
以下利用查询单词 “schema” 为例说明上述查询动作。首先，利用此单词的前两个字母 “sc” 来对二次索引数据 15 进行查询。透过公式(1)，可



知其二次索引值储存于二次索引数据 15 的第  $(19-1) \times 27 + 3 = 489$  个位置。因此，在二次索引数据 15 的第 489 个和第 490 个位置上，可以找到二次索引值 I1 和 I2。再利用二次索引值 I1 和 I2，可以从单词索引数据 20 找到对应的单词索引值 R1 和 R2。再以单词索引值 R1 和 R2，以折半查询方式找出对应于“schema”的单词索引值，接着便可以查询出其对应的相关内容。

从上述说明可知，从于在本实施例中的查询动作是先利用二次索引数据来决定出折半查询范围，因此降低了执行折半查询/对比的次数，不仅可以缩短查询时间，同时减少对单词数据库(一般为压缩形式)进行解压缩的处理次数，提高整体单词查询的性能。

本实施例的查询索引结构是非常适合于一般电子字典中动态的查询方式，亦即随着用户逐次输入单词的各字母，能够快速查询出与目前输入部分者相同或最接近的单词的相关内容。图 10 表示本实施例中执行动态单词查询的处理流程图。首先，系统依序接收用户所输入的单词字母(S1)。当输入单词的字母数不大于两个(有序序列的预定字母数)时(S2)，则可以直接利用二次索引数据 15 来快速完成查询。先利用目前所输入的单词，计算出单词输入字母在二次索引数据 15 的位置，并且取得其所对应的二次索引值 I1(S3)。接着再利用二次索引值 I1 和下一个项目的二次索引值 I2，找出其对应的单词索引值 R1、R2，用以确定折半查询范围为  $R1 \sim R2$ (S4)。接着则利用二次索引值 I1 所得到的待查单词的单词索引值 R1(S5)，查询单词数据库 30 和单词相关内容数据库 50(S6)，找出目前输入部分所对应的最相似单词及其相关内容，并且在电子字典的显示器(未图示)上则显示出所查询出的最相似单词及其相关内容(S7)，完成阶段性的查询工作。当用户输入的字母数超过两个以上时(S2)，则是以先前所决定的单词索引值  $R1 \sim R2$  范围间进行折半查询，找出最接近于目前输入单词部分的待查单词的单词索引值 R3，并且以 R3 更新上述的查询范围中的 R1(S8)。同样的，利用单词索引值 R1(已更新)来查询单词数据库 30 和单词相关内容数据库 50(S6)，找出目前输入部分所对



应的最相似单词以及其相关内容，并且在电子字典的显示幕上显示所查询出的最相似单词及其相关内容(S7)。上述动态查询动作会在用户结束输入或者是已查询到所需要单词时终止。

配合图 3 ~ 图 6，利用用户动态输入查询单词“schema”为范例，说明图 10 所示的动态查询动作。在图 3 中，用户输入部分查询单词“s”，从于尚未超过两个字母，所以利用公式(1)计算出在二次索引数据 15 中的对应储存位置：

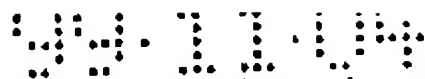
$$\text{储存位置} = (\text{字母“s”顺序编号} - 1) \times 27 + 0 = (19-1) \times 27 + 0 + 486$$

因此，在二次索引数据 15 的第 486 项储存位置中可得到以“s”为首的  
10 第一个单词的二次索引值 I1，接着利用二次索引值 I1，可得到在单词索引数据 20 的第 I1 项中所储存的单词索引值 R1，其表示以“s”为首的  
第一个单词在单词数据库 30 中的储存位置。同时可以在第 I1 + 1 项、第 I1 + 2 项  
等等得到其后续单词的单词索引值。在图 3 的左半部即显示目前最接近单词  
(即“s”)以及其后续单词。同时，利用单词索引值 R1 可以透过单词相关内  
15 容索引数据 40 以及单词相关内容数据库 50 查询出对应于此最接近单词(即  
“s”)的相关内容，如图显示于右半部。

在图 4 中，用户输入了部分查询单词“sc”，同样未超过两个字母，所以再利用公式(1)计算出在二次索引数据 15 中的对应储存位置：

$$\begin{aligned} \text{储存位置} &= (\text{字母“s”顺序编号} - 1) \times 27 + (\text{字母“c”顺序编号}) \\ 20 \quad &= (19-1) \times 27 + 3 = 489 \end{aligned}$$

因此，在二次索引数据 15 的第 489 项储存位置中可得到以“sc”为首的  
第一个单词的二次索引值 I1，接着利用二次索引值 I1，可得到在单词索引  
引数据 20 的第 I1 项中所储存的单词索引值 R1，其表示以“sc”为首的  
第一个单词在单词数据库 30 中的储存位置。同样的，利用单词索引值 R1 可以  
25 透过单词相关内容索引数据 40 以及单词相关内容数据库 50 查询出对应于此  
最接近单词(即“sc”)的相关内容，另外，在处理此阶段的输入单词时，可



以建立后续输入阶段所需要的折半查询范围。在上述处理中所得到的单词索引值 R1 为此范围的一端，其对应于以“sc”为首的第一个单词。接着，从二次索引数据 15 的第 490 项储存位置中得到其下一个的二次索引值 I2，再从单词索引数据 20 找到第 I2 项中所储存的单词索引值 R2，其为范围的另一端，对应于以“sd”为首的第一个单词。亦即，无论用户后续所输入的字母为何，此单词所对应的单词索引值会在 R1(对应于“sa”为首的第一个单词)和 R2(对应于“sd”为首的第一个单词)之间。

接着在图 5 中，用户输入了部分查询单词“sch”，已经超过两个字母，因此可以直接利用先前所建立的折半查询范围 R1 ~ R2 来进行查询。当找到以“sch”为首的第一个单词以及其单词索引值 R3 时，则显示所查询到的最接近单词(即“SCH”)和后续单词以及其相关内容。另外，利用最接近单词的单词索引值 R3 来取代上述的折半查询范围值 R1，亦即无论用户后续所输入的字母为何，此单词所对应的单词索引值会在更新后的 R1(对应于“sch”为首的第一个单词)和 R2(对应于“sd”为首的第一个单词)之间。同样的处理适用于用户后续输入的各字母，直到如图 6 所示，用户输入“schema”之后，便完成所有查询的动作。

本实施例的索引数据结构以及动态查询方法，其优点便在于可以节省大量反复折半查询的次数。表 1 是用以比较现有技术和本发明的平均查询次数。一般标准折半查询的次数为  $\log_2(N)$ ，其中 N 表示待查询总项目。以字典单词数 50000 ~ 100000 为例，现有技术平均需要 15.6 到 16.6 次的查询动作。而在本发明中，从于利用二次索引数据先将所有单词区分为 702 类(以两个字母来分类)，所以平均只需要 6.1 到 7.1 次查询动作，便可以找到所需要单词。

表 1

字典单词数	现有技术(次)	本发明(次)
50000	$\log_2 50000 = 15.6$	$\log_2(50000/702) = 6.1$
100000	$\log_2 100000 = 16.6$	$\log_2(100000/702) = 7.1$



如果以 50000 字的字典为例分析逐次输入各单词字母的平均查询次数, 则如表 2 所示。在表 2 中, 输入字母数为累计次数, 更可以看出本发明在减少查询次数上的效果。

表 2

输入字母数	现有技术(次)	本发明(次)
1	$\log_2 50000 = 15.6$	1
2	$2 \times \log_2 50000 = 31.2$	2
5	$5 \times \log_2 50000 = 78$	$3 \times \log_2(50000/702) + 2 = 20.3$
10	$10 \times \log_2 50000 = 156$	$6 \times \log_2(50000/702) + 2 = 38.6$

- 5 本发明虽以一优选实施例公开如上, 然其并非用以限定本发明, 本领域技术人员, 在不脱离本发明的精神和范围内, 可做某些改动和替代, 因此本发明的保护范围当从后附的权利要求所界定。



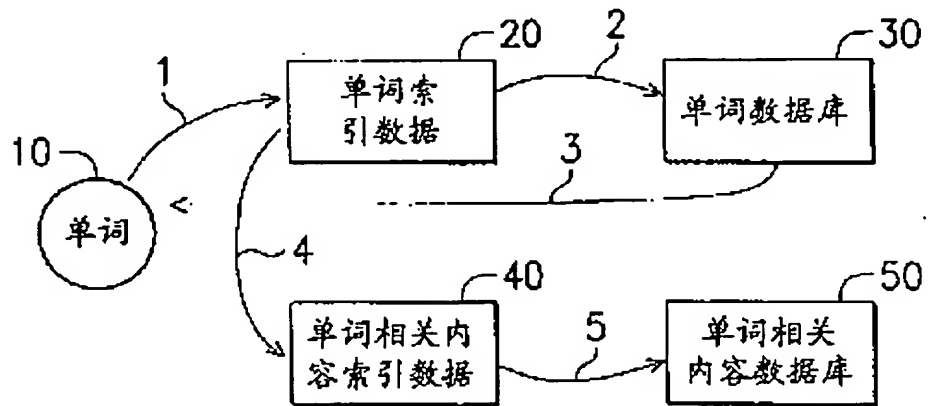


图 1

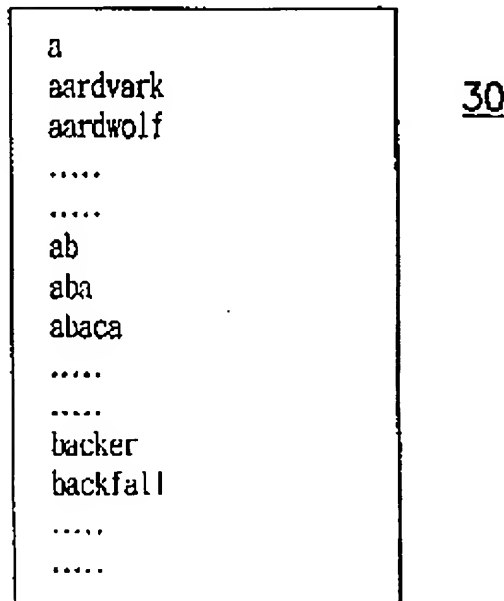


图 2

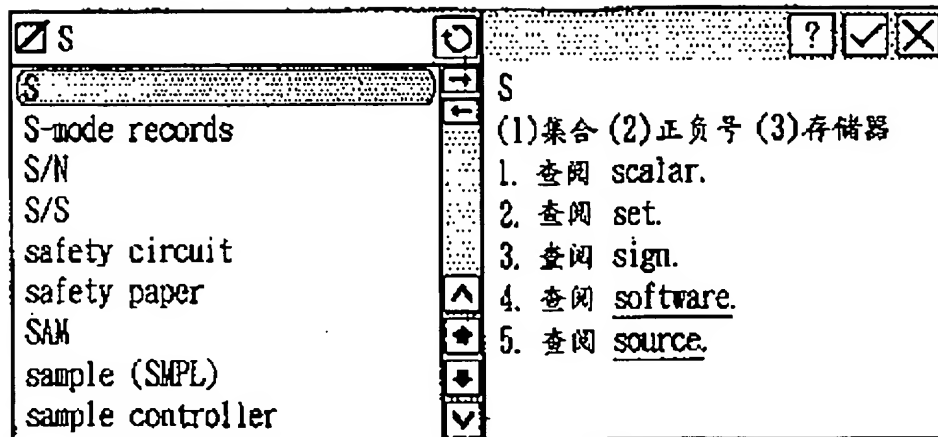


图 3

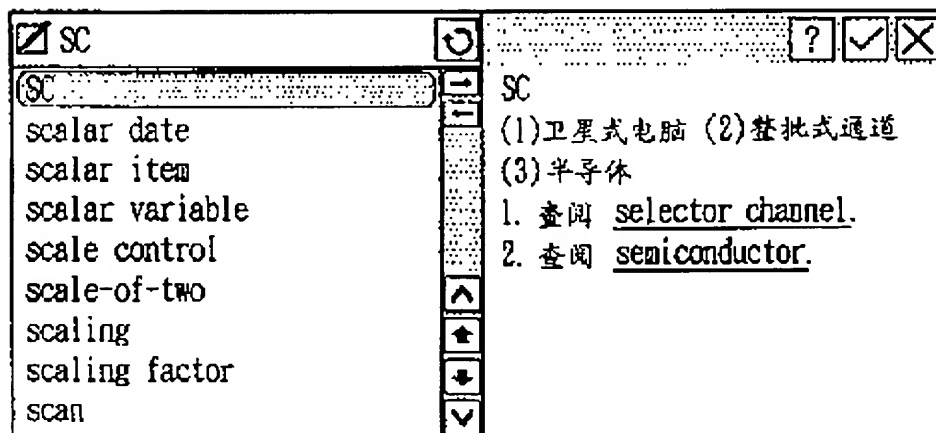
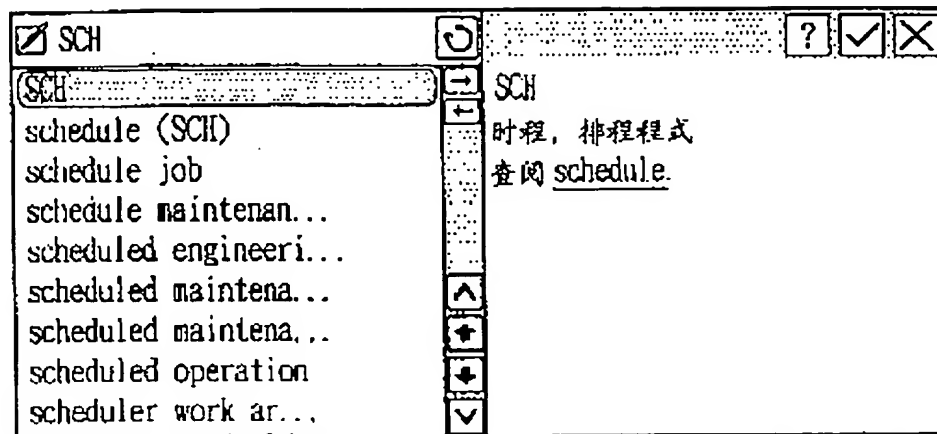


图 4



The screenshot shows a file explorer window titled 'Index of'. The left pane displays a list of files, with 'schema' selected. The right pane shows the contents of the selected file, which is a text file. The text in the right pane is as follows:

```

schema
描述语言
根据整个数据库观念模型说明
逻辑及实际的结构

```

3

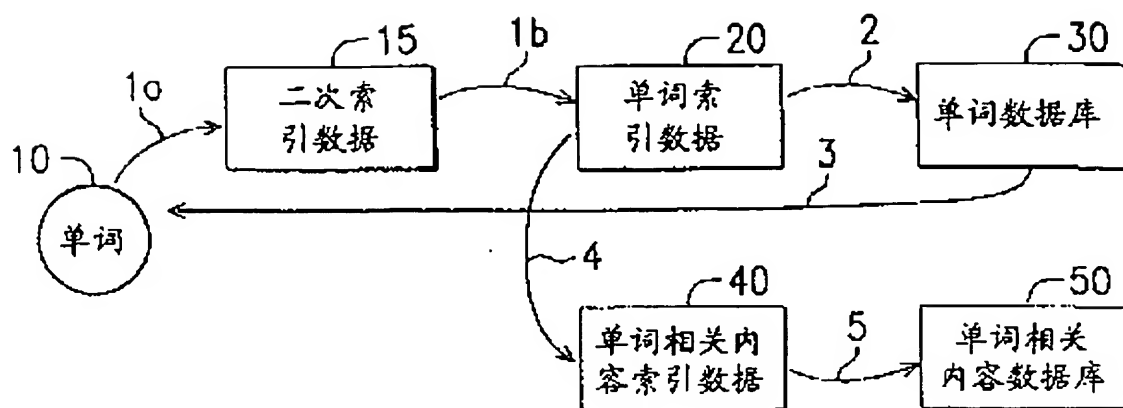
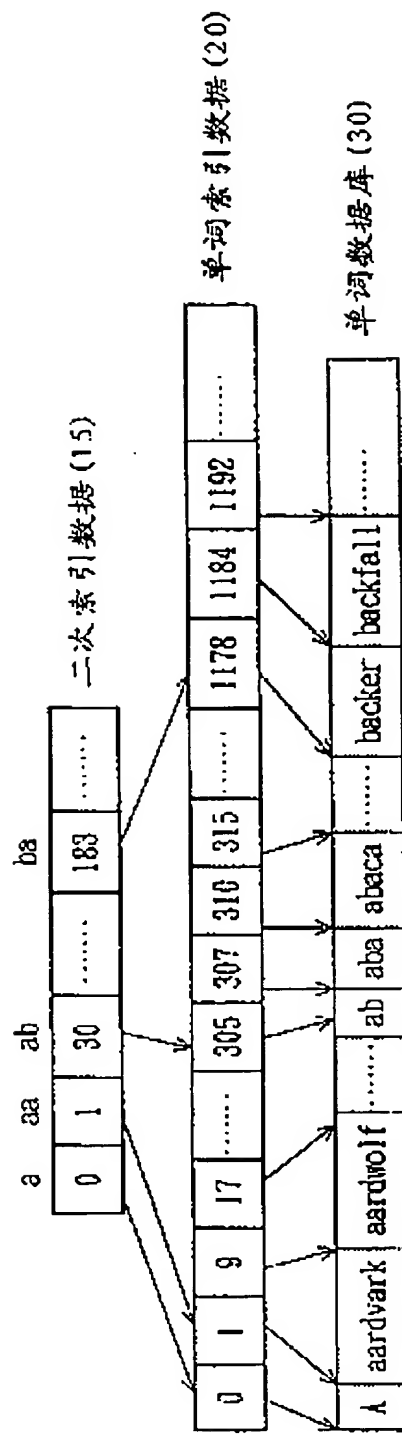


图 7

15b	15a	
a	0	
aa	1	
ab	30	
ac	47	
...	...	
...	...	
b	183	
ba	...	
bb	...	
bc	...	
...	...	
...	...	
c	...	
ca	...	
cb	...	
cc	...	
...	...	
...	...	
zy	...	
zz	...	

15

图 8



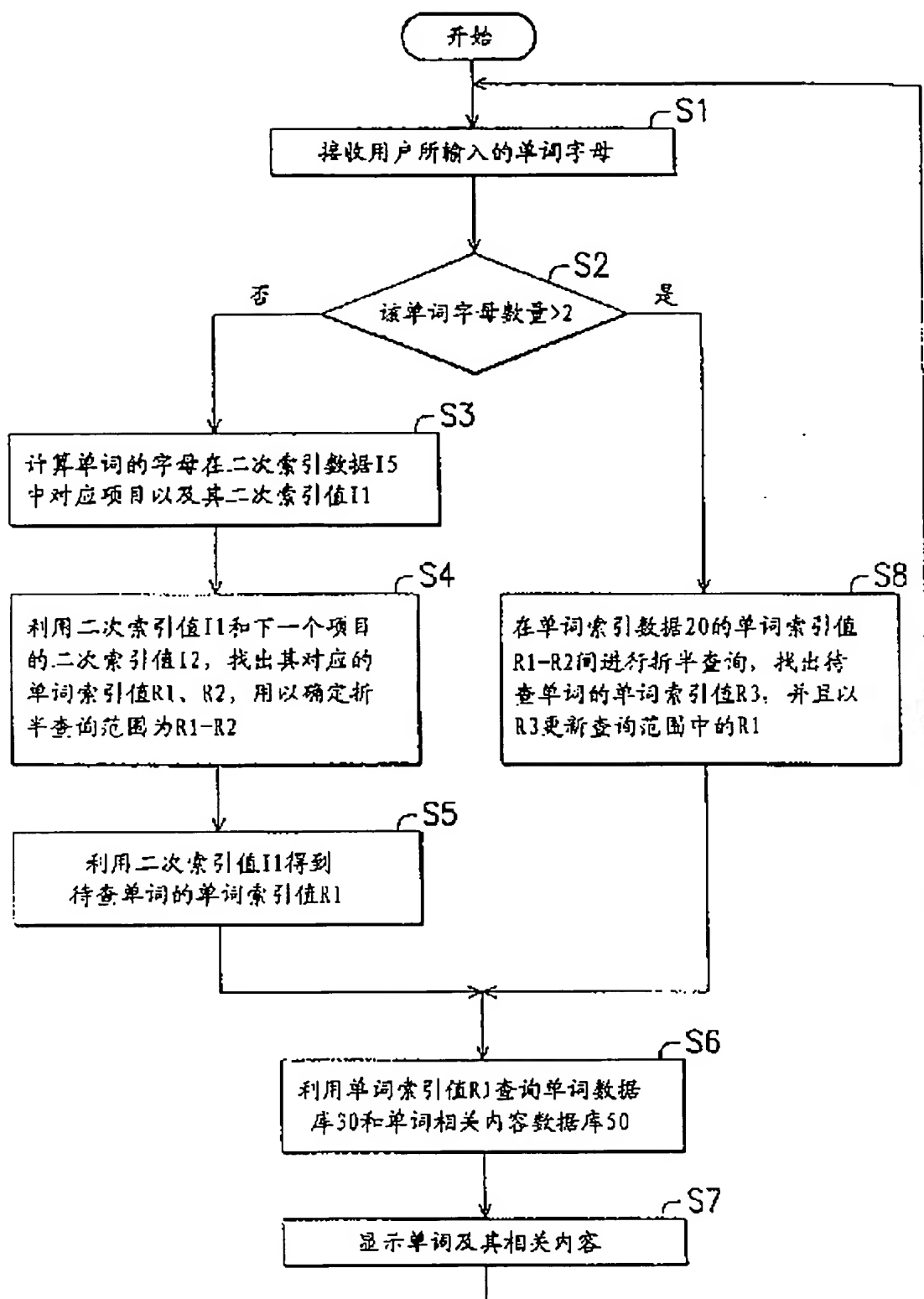


图 10

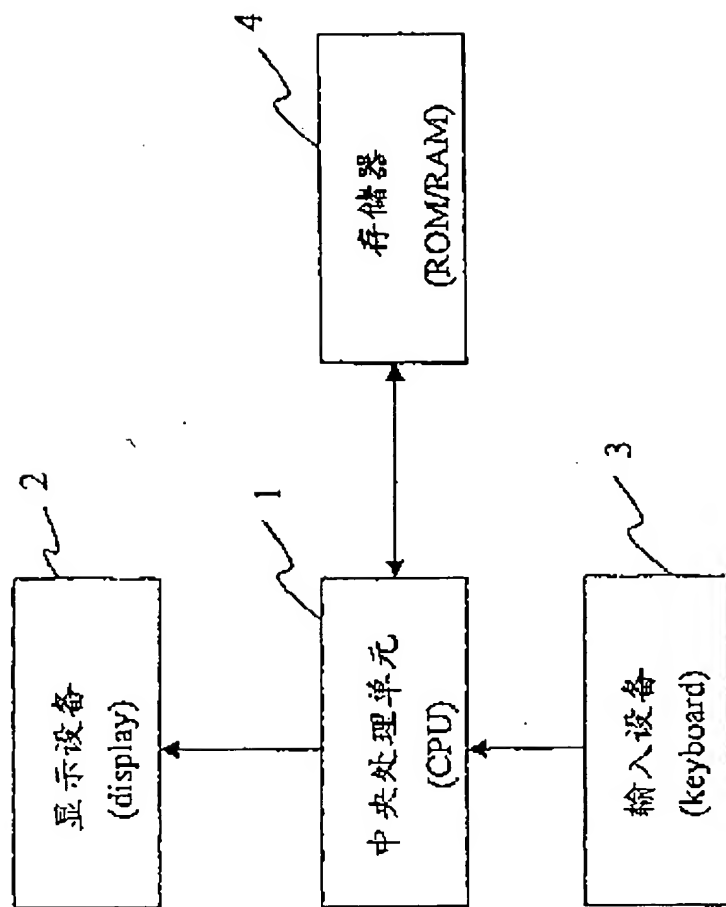


图 11